Week 5 Report: Hypothesis Testing 2

Data Analytics & Machine Learning Research Lab

Vo Nguyen Le Duy

Tran Minh Thien

26th Febuary 2025

Contents

1 Hypothesis testing x

1.1 Problem

- Given an array $\mathbf{X}$ $[x1, ..., x5]^T$. Perform Hypothesis testing to check whether an element in $\mathbf{X}$ equal 0 or not (in this case is **second value** $\sim \mathbf{x_2}$)
- The **Null hypothesis** ($\mathbf{H_0}$) is:

$$x_2 = 0$$

- The **Alternative hypothesis** ($\mathbf{H_1}$) is:

$$x_2 \neq 0$$

1.2 Analysis

1. Firstly, Test-Statistic is $\mathbf{x_2}$: $T = x_2 = \eta^T X = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 \end{bmatrix} \cdot \begin{bmatrix} x1 \\ x2 \\ x3 \\ x4 \\ x5 \end{bmatrix} \rightarrow \eta = \begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}$
2. Secondly, distribution of T is belong to η: $T = \eta^T.y \sim (\eta^T \mu, \eta^T \sigma^2 \eta)$
   X is generate with Normal Distribution $\rightarrow \sigma^2 = 1 \rightarrow T \sim (\eta^T \mu, \eta^T \eta)$

1.3 Graph

- Histogram of P_value

$$FPR \text{ (False Positive Rate)} = 5.14\% \;\approx\; 257/5000$$

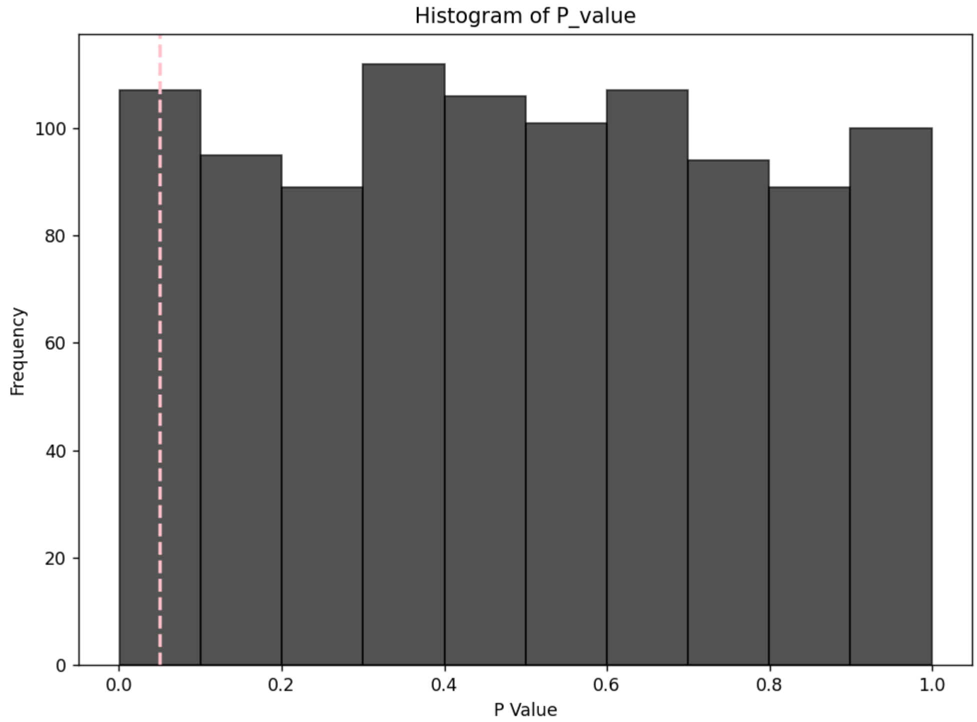

Figure 1: P_Value

2 Hypothesis testing the largest x

2.1 Problem

- Identical to Problem 1 but instead of testing a permanent element. Testing the largest element of each test.
- The **Null hypothesis** ($\mathbf{H_0}$) is:
$$x_2 = 0$$
- The **Alternative hypothesis** ($\mathbf{H_1}$) is:
$$x_2 \neq 0$$

2.2 Analysis

1. j = index of the largest element = **argmax**(**X**)

2. Test-Statistic is $\mathbf{x_j}$: $T = x_j = \eta^T X = \begin{bmatrix} x_{1,j} & x_{2,j} & x_{3,j} & x_{4,j} & x_{5,j} \end{bmatrix} \cdot \begin{bmatrix} x1 \\ x2 \\ x3 \\ x4 \\ x5 \end{bmatrix}$

$$\rightarrow \eta = \begin{bmatrix} x_{1,j} \\ x_{2,j} \\ x_{3,j} \\ x_{4,j} \\ x_{5,j} \end{bmatrix} \leftarrow x_{i,j} \begin{cases} 1, & \text{if } i = j \\ 0, & \text{if } i \neq j \end{cases}$$

3. The distribution of T belongs to η: $T = \eta^T.y \sim (\eta^T \mu, \eta^T \sigma^2 \eta)$ X is generated with Normal Distribution $\rightarrow \sigma^2 = 1 \rightarrow T \sim (\eta^T \mu, \eta^T \eta)$

2.3 Graph

- Histogram of P_value

$$FPR \text{ (False Positive Rate)} = 13.7\% \;\approx\; 685/5000$$

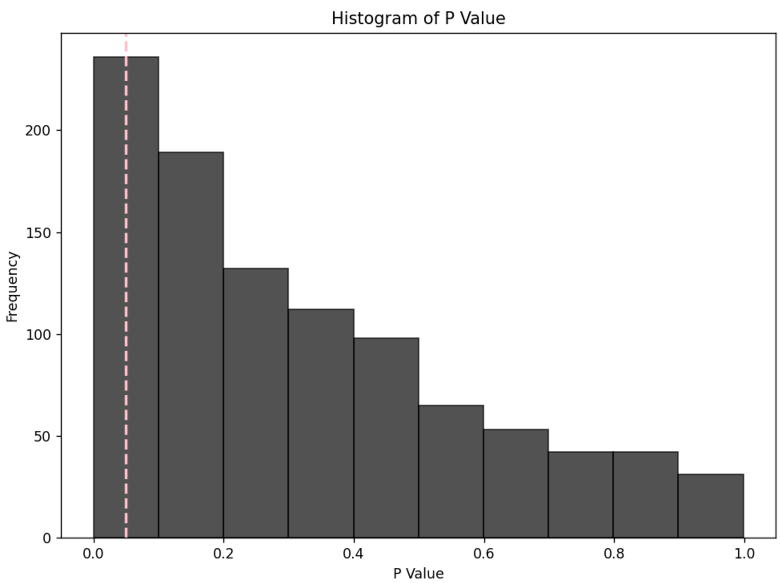

Figure 2: P_Value

3 Hypothesis testing for Linear Regression with $\hat{\beta}$ selected randomly

3.1 Problem

- Given a Linear Regression with p (p = 5) features. Randomly select 1 of the feature and perform hypothesis testing to check whether $\hat{\beta}_j$ equals 0 or not.
- The **Null hypothesis** ($\mathbf{H_0}$) is:

$$\hat{\beta}_j = 0$$

- The **Alternative hypothesis** ($\mathbf{H_1}$) is:

$$\hat{\beta}_j \neq 0$$

3.2 Analysis

1. Firstly, j = random(0, p - 1) $\rightarrow$ X$_j$ $= \eta X$

$$\rightarrow \eta = \begin{bmatrix} x_{1,j} \\ x_{2,j} \\ x_{3,j} \\ x_{4,j} \\ x_{5,j} \end{bmatrix} \leftarrow x_{i,j} \begin{cases} 1, & \text{if } i = j \\ 0, & \text{if } i \neq j \end{cases}$$

2. Secondly, Test-Statistic is $\hat{\beta}_j$: $T = \hat{\beta}_j = \eta_j^T y = (X^T X)^{\dagger} X^T y$
   $\rightarrow \eta_j^T = (\mathbf{X^T X})^{\dagger} \mathbf{X^T}$

3. Thirdly, distribution of T is belong to η_j: $T = \eta_j^T.y \sim (\eta_j^T \mu, \eta_j^T \sigma^2 \eta)_j$
   X is generate with Normal Distribution $\rightarrow \sigma^2 = 1 \rightarrow T \sim (\eta_j^T \mu, \eta_j^T \eta_j)$

3.3 Graph

- Histogram of **P_value**

$$FPR \text{ (False Positive Rate)} = 5.64\% \;\approx\; 282/5000$$

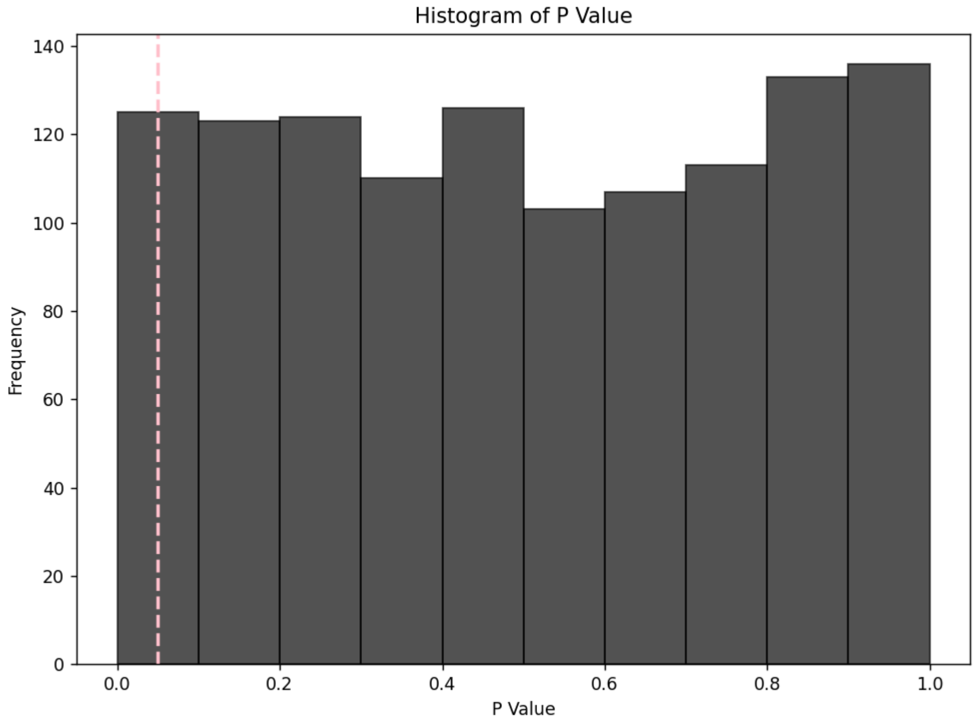

Figure 3: P_value

4 Hypothesis testing for Linear Regression with the largest $\hat{\beta}$

4.1 Problem

- Given a Linear Regression with 5 features. Select the largest **Weight** and perform hypothesis testing to check whether $\hat{\beta}_j$ equals 0 or not.

- The **Null hypothesis** ($\mathbf{H_0}$) is:

$$\hat{\beta}_j = 0$$

- The **Alternative hypothesis** ($\mathbf{H_1}$) is:

$$\hat{\beta}_j \neq 0$$

4.2 Analysis

1. Firstly, j = argmax($\hat{\beta}$) $\rightarrow$ X$_j$ = ηX

$$\rightarrow \eta = \begin{bmatrix} x_{1,j} \\ x_{2,j} \\ x_{3,j} \\ x_{4,j} \\ x_{5,j} \end{bmatrix} \leftarrow x_{i,j} \begin{cases} 1, & \text{if } i = j \\ 0, & \text{if } i \neq j \end{cases}$$

2. Secondly, Test-Statistic is $\hat{\beta}_j$: $T = \hat{\beta}_j = \eta^T y = (X^T X)^{\dagger} X^T y$
   $\rightarrow \eta^T = (\mathbf{X^T X})^{\dagger}\mathbf{X^T}$

3. Thirdly, distribution of T is belong to η: $T = \eta^T.y \sim (\eta^T\mu, \eta^T\sigma^2\eta)$
   X is generate with Normal Distribution $\rightarrow \sigma^2 = 1 \rightarrow T \sim (\eta^T\mu, \eta^T\eta)$

4.3 Graph

- Histogram of **P_value**

$$FPR \text{ (False Positive Rate)} = 13.08\% \;\approx\; 654/5000$$

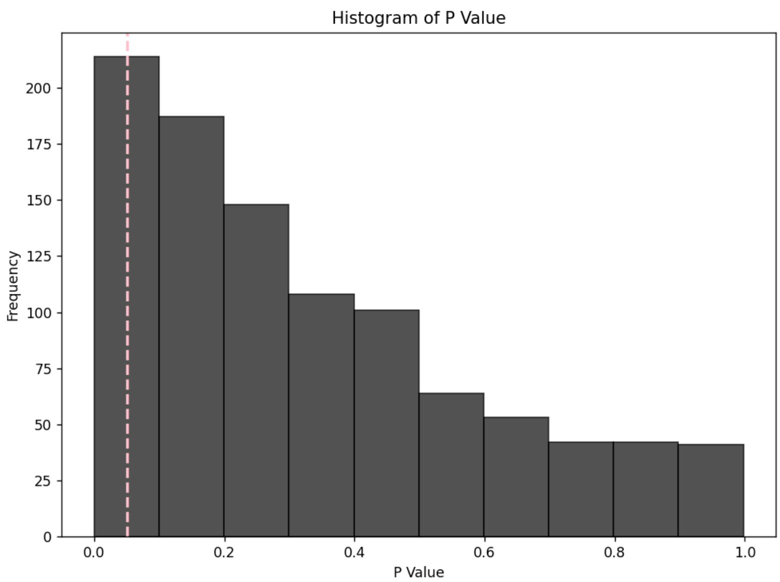

Figure 4: P_Value